Amendments to Claims

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims

1. (previously presented) A method for segmenting compound words in an unrestricted natural-language input, the method comprising:

      receiving a natural-language input consisting of a plurality of characters;

      constructing a set of breakpoints in the natural-language input;

      combining a probability that characters preceding each breakpoint end a word and a probability that characters following the breakpoint start a word to assign weights to the breakpoints in the natural-language input;

      traversing substrings of the natural-language input in an order determined by the weights assigned to the breakpoints;

      identifying a plurality of linkable components by the traversal of substrings wherein a linkable component is identified by locating the component in a lexicon; and

      returning a segmented string consisting of a plurality of linkable components spanning the natural-language input, wherein the segmented string is interpreted as a compound word.

2. (original) The method of claim 1, further including the step of analyzing a chart of the linkable components in the case that the segmented string cannot be constructed and returning an unsegmented string interpretable as a partial analysis of a compound word.

.

3. (previously presented) An apparatus for segmenting compound words in a natural-language input, the apparatus comprising:

      a startpoint probability matrix;

      an endpoint probability matrix;

      a probabilistic breakpoint analyzer coupled to the startpoint probability matrix, the endpoint probability matrix and the natural-language input, the probabilistic breakpoint analyzer being operative to generate a breakpoint-annotated input from the natural-language input; and

2

a probabilistic breakpoint processor coupled to the probabilistic breakpoint analyzer, the probabilistic breakpoint processor being operative to generate a segmented string for the compound words in the natural-language input in response to the breakpoint-annotated input.

4. (original) The apparatus of claim 3, further comprising a word-boundary analyzer coupled to a lexicon and a memory unit, the word-boundary analyzer being operative to generate the startpoint probability matrix and the endpoint probability matrix.

5. (original) The apparatus of claim 3, wherein the probabilistic breakpoint processor comprises:

a lexicon;

a chart; and

a breakpoint-delimited substring tester coupled to the lexicon and the chart, the substring tester being operative to receive the breakpoint-annotated input and generate a segmented string in response thereto.

6. (original) The apparatus of claim 3, wherein the probabilistic breakpoint processor is an augmented probabilistic breakpoint processor comprising:

a lexicon;

a chart;

an augmented breakpoint-delimited substring tester coupled to the chart and the lexicon, the substring tester being operative to identify a plurality of linkable components; and

a chart analyzer coupled to the substring tester and the chart, the chart analyzer being operative to generate the segmented string.

7. (original) The apparatus of claim 6, wherein the augmented breakpoint-delimited substring tester generates one of:

the segmented string; and

a failure signal.

8. (original) The apparatus of claim 7, wherein the chart analyzer is coupled to receive the failure signal from the augmented breakpoint-delimited substring tester.

9. (original) The apparatus of claim 3, wherein the apparatus is configured as a computer readable program code run on a computer usable medium.

10. (cancelled)

11. (previously presented) A method for segmenting compound words in an unrestricted natural-language input, the method comprising:

 receiving a natural-language input consisting of a plurality of characters;

 constructing a set of breakpoints in the natural-language input;

 combining weights of trigraph contexts that precede and follow each breakpoint to assign a weight to the breakpoint in the natural-language input;

 traversing substrings of the natural-language input in an order determined by the weights assigned to the breakpoints;

 identifying a plurality of linkable components by the traversal of substrings wherein a linkable component is identified by locating the component in a lexicon; and

 returning a segmented string consisting of a plurality of linkable components spanning the natural-language input, wherein the segmented string is interpreted as a compound word.

12. (previously presented) A method for segmenting compound words in an unrestricted natural-language input, the method comprising:

 receiving a natural-language input consisting of a plurality of characters;

 constructing a set of breakpoints in the natural-language input;

 combining weights of bigraph contexts that precede and follow each breakpoint to assign a weight to the breakpoint in the natural-language input;

 traversing substrings of the natural-language input in an order determined by the weights assigned to the breakpoints;

 identifying a plurality of linkable components by the traversal of substrings wherein a linkable component is identified by locating the component in a lexicon; and

returning a segmented string consisting of a plurality of linkable components spanning the natural-language input, wherein the segmented string is interpreted as a compound word.

13. (previously presented) A method for segmenting compound words in an unrestricted natural-language input, the method comprising:

receiving a natural-language input consisting of a plurality of characters;

constructing a set of breakpoints in the natural-language input;

combining weights of tetragraph contexts that precede and follow each breakpoint to assign a weight to the breakpoint in the natural-language input;

traversing substrings of the natural-language input in an order determined by the weights assigned to the breakpoints;

identifying a plurality of linkable components by the traversal of substrings wherein a linkable component is identified by locating the component in a lexicon; and

returning a segmented string consisting of a plurality of linkable components spanning the natural-language input, wherein the segmented string is interpreted as a compound word.

14. (previously presented) A method for segmenting compound words in an unrestricted natural-language input, the method comprising:

receiving a natural-language input consisting of a plurality of characters;

constructing a set of breakpoints in the natural-language input;

combining weights of contexts of one length that precede each breakpoint and of contexts of a different length that follow the breakpoint to assign a weight to the breakpoint in the natural-language input;

traversing substrings of the natural-language input in an order determined by the weights assigned to the breakpoints;

identifying a plurality of linkable components by the traversal of substrings wherein a linkable component is identified by locating the component in a lexicon; and

returning a segmented string consisting of a plurality of linkable components spanning the natural-language input, wherein the segmented string is interpreted as a compound word.

5

15. (previously presented) A method for segmenting compound words in an unrestricted natural-language input, the method comprising:

      receiving a natural-language input consisting of a plurality of characters;

      constructing a set of breakpoints in the natural-language;

      weighting weights of a plurality of contexts of different lengths that precede and follow each breakpoint to assign a weight to the breakpoint in the natural-language input;

      traversing substrings of the natural-language input in an order determined by the weights assigned to the breakpoints;

      identifying a plurality of linkable components by the traversal of substrings wherein a linkable component is identified by locating the component in a lexicon; and

      returning a segmented string consisting of a plurality of linkable components spanning the natural-language input, wherein the segmented string is interpreted as a compound word.